Supplementary Online Content

Shear MK, Reynolds III CF, Simon NM, et al. Optimizing treatment of complicated grief: a randomized clinical trial. *JAMA Psychiatry*. Published online June 8, 2016. doi:10.1001/jamapsychiatry.2016.0892.

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix 1. Details of the main statistical analyses

We used Inverse Probability Weighting (IPW), a standard strategy to account for missing assessment data[21-24], to analyze the primary outcomes and secondary outcomes under the intention-to-treat (ITT) principle including all randomized participants. IPW is a two-stage procedure that first determines predictors of assessment completion. In this first stage, a logistic regression model is used to compute the probability of completing a follow up assessment at a given week (week 12 or week 20) using the following as predictors for each subject: treatment assignment, site, baseline demographics, baseline comorbidities (MDD and PTSD), baseline symptom severities, as well as interactions between treatment assignment and the other variables. Each assessment completer is then assigned a weight variable based on the inverse probability of her/his predicted completion probability. Thus, more weight is assigned to completers who represent those who did not complete the assessment. Robust variance estimators were computed to account for the uncertainties due to estimating those weights. In a second stage, using weights constructed in a first stage, a weighted logistic regression, controlling for randomization stratification variables and baseline covariates found to be imbalanced across treatment arms, was used to estimate the CGI-Improvement response rate (primary outcome) at week 12 or week 20 and compare across treatment arms for each aim. A similar weighted linear regression was used to estimate the mean continuous secondary outcomes at week 12 or 20 (self-report ratings of ICG, WSAS, GRAQ, and QIDS-SR_{16}) and compare across arms.

Table 1. IPW adjusted response rate (primary outcome) in each treatment arm[+]

	CIT	PLA	CIT + CGT	PLA + CGT	RR (Relative Risk)	95% CI RR	p-value	NNT	95% CI NNT
Aim 1: CIT vs. PLA (week 12)	45.90 %	37.90 %			1.21	(0.81, 1.81)	0.35	12.4	(NNTH 11.8 to ∞ to NNTB 4.1)[#]
Aim 2: CIT + CGT vs. PLA + CGT (week 20)			83.73 %	82.54 %	1.01	(0.88, 1.17)	0.84	84.0	(NNTH 9.4 to ∞ to NNTB 7.7)
Aim 3: CIT + CGT vs. CIT (week 20)	69.33 %		83.73 %		1.21	(1.00, 1.46)	0.05*	6.9	(3.5, 207.6)
Secondary aim: PLA + CGT vs. PLA (week 20)		54.81 %		82.54 %	1.51	(1.16, 1.95)	0.002*	3.6	(2.3, 8.1)

[+]: Results obtained by weighted logistic regressions of the primary outcome measured at the primary time points (week 12 or week 20), controlling for site, baseline MDD, ethnicity, and with IPW to account for missing assessment data.
[#]: NNTH: numbers needed to harm; NNTB: numbers needed to benefit.
*: Significant at a significance level of 0.05.

As sensitivity analyses, generalized linear mixed effects model with random intercepts and a logit link function was used for the binary CGI-Improvement responder status, where all available assessments were included in the analysis. To be consistent with the ITT IPW cross-sectional analysis, we report relative risk obtained by taking the ratio of the response rates estimated from the model, instead of reporting the odds ratio. Linear mixed effects model was used for the continuous secondary outcomes (ICG, WSAS, GRAQ, QIDS-SR_{16}), controlling for the same set of variables as the primary analyses, and using the actual assessment time instead of nominal time. The estimated parameters from the model were then used to provide an estimate of the adjusted mean difference in the outcome at the primary time point (week12 or 20) for each aim. A quadratic time effect was included for QIDS-SR_{16} to fit the observed data.

Site is adjusted as a fixed effect in the IPW regression analyses. Due to the small number of participating sites in our study, it is not feasible to include it as a random effect. Thus, it is not appropriate to generalize our results to the population beyond the participating sites.

eAppendix 2. Results of the moderator analyses on the primary outcome

A pre-planned regression analysis of baseline MDD as a moderator examined treatment by MDD interaction by testing the significance of interaction effect in the weighted logistic and weighted linear model. In Table 2, we report the IPW adjusted response rate in participants with and without MDD.

Table 2.

		IPW Adjusted Response Rates							
		CIT	PLA	CIT + CGT	PLA + CGT	RR	95% CI RR	p-value	interaction p-value
Aim 1: CIT vs. PLA (week 12)	MDD	45.74 %	30.77 %			1.5	(0.88, 2.58)	0.14	0.31
	No MDD	46.40 %	49.18 %			0.97	(0.52, 1.83)	0.93	
Aim 2: CIT + CGT vs. PLA + CGT	MDD			84.96 %	77.76 %	1.09	(0.91, 1.31)	0.35	0.18
	No MDD			84.31 %	93.01 %	0.9	(0.72, 1.12)	0.34	
Aim 3: CIT + CGT vs. CIT	MDD	72.11 %		84.96 %		1.18	(0.96, 1.45)	0.14	0.68
	No MDD	65.04 %		84.31 %		1.29	(0.89, 1.86)	0.17	

eAppendix 3. Results of the longitudinal analyses on the primary and secondary outcomes

Longitudinal analyses results for the primary outcome (response rate) mirrored the IPW cross-sectional results. There was no significant difference in response rate between CIT and PLA (RR=1.25, p=0.48), and no significant difference between CIT + CGT and PLA + CGT (RR=1.04, p=0.43). There was a significant difference between CIT + CGT and CIT (RR=1.24, p=0.01), and a significant difference between PLA + CGT and PLA (RR=1.51, p=0.003).

Longitudinal analyses for CG symptom and impairment measures using all available assessments on the ITT sample were also consistent with the IPW cross-sectional analyses. CIT and PLA did not differ at week 12 on self-report ratings of CG symptom severity (ICG: adjusted mean difference = -1.22, p=0.41) grief-related impairment (WSAS: adjusted mean difference = -0.66, p=0.56) or grief-related avoidance (GRAQ: adjusted mean difference = -0.17, p=.89). Similarly, CIT + CGT and PLA + CGT did not differ on any of these measures at week 20. Also mirroring the main outcome results and IPW cross-sectional results, CIT + CGT was associated with significantly greater change in self-report ratings than CIT alone (ICG adjusted mean difference = -5.37, p=0.0005; WSAS adjusted mean difference=-3.06, p=0.01; GRAQ adjusted mean difference = -3.77, p=0.003) and PLA + CGT was associated with significantly greater change in self-report ratings than PLA alone (ICG adjusted mean difference = -8.01, p=0.0001; WSAS adjusted mean difference=-5.76, p=0.0004; GRAQ adjusted mean difference = -5.74, p=0.0001). Longitudinal analyses of depressive symptom severity (QIDS-SR_{16}) showed no significant difference at week 12 between CIT and PLA (mean difference = -0.92, p =

0.09). In contrast, there was a marginally significant difference at week 20 between CIT + CGT and PLA + CGT (mean difference = -1.24, p=0.06). A significant difference was also observed between CIT + CGT and CIT (mean difference = -2.03, p=0.002), and between PLA + CGT and PLA (mean difference = -2.6, p=0.0001).